IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 5, 2019, accepted October 25, 2019, date of publication November 6, 2019, date of current version November 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951762

Isolated Pulmonary Nodules Characteristics Detection Based on CT Images

SHI QIU[1], QIANG GUO[2], DONGMEI ZHOU[3], YI JIN[4], TAO ZHOU[5], AND ZHEN'AN HE[6]

[1]Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
[2]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China
[3]School of Information Science and Technology, Chengdu University of Technology, Chengdu 610059, China
[4]School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
[5]School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China
[6]Shaanxi Institute of Medical Device Quality Supervision and Inspection, Xi'an 712046, China

Corresponding author: Qiang Guo (guoqiang@sdufe.edu.cn)

This work was supported by in part by the National Natural Science Foundation of China under Grant 61873145, in part by the Natural Science Foundation of Shandong Province for Excellent Young Scholars under Grant ZR2017JL029, in part by the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions, and in part by the National Natural Science Foundation of China under Grant 61561040.

ABSTRACT Pulmonary nodules are the main pathological changes of the lung. Malignant pulmonary nodules will be transformed into lung cancer, which is a serious threat to human health and life. Therefore, the detection of pulmonary nodules is of great significance to save lives. However, in the face of a large number of lung CT image sequences, doctors need to spend a lot of time and energy, and in the detection process will inevitably produce the problem of false detection and missed detection. Therefore, it is very necessary for computer-aided doctors to detect pulmonary nodules. It is difficult to segment pulmonary nodules accurately and recognize the characteristics of pulmonary nodules in CT images. A complete set of semi-automatic lung nodule extraction and feature identification system is established, which is in line with the doctor's diagnosis process. A segmentation algorithm of pulmonary nodules based on regional statistical information is proposed to extract pulmonary nodules accurately. This is the first time that dynamic time warping algorithm is applied in the field of image processing, focusing on the lung nodule boundary. On this basis, the recursive graph visualization model is established to realize the visualization of boundary similarity. Finally, in order to accurately identify the characteristics of pulmonary nodules, a video similarity distance discrimination system is introduced to quantify the similarity between the nodules to be examined and the pulmonary nodules in the database. The experimental results show that the algorithm can accurately identify the normal shape, lobulated shape and lobulated shape of pulmonary nodules. The average processing speed is 0.58s/nodule. To some extent, it can reduce the misdiagnosis caused by experience and fatigue.

INDEX TERMS Pulmonary nodules, time series, recurrence plot, characteristics, computer detection.

I. INTRODUCTION

Lung cancer is a common malignant tumor with high morbidity, rapid growth and other features, and it has huge obstacle and hidden danger to human life and health [1], [2]. The isolated pulmonary nodule which divided into benign and malignant [3] is the most common manifestations of lung cancer in the early stages. Accordingly, early confirmation and treatment of pulmonary nodules are of great significance to saving human lives.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongtao Hao.

CT scanning image is based on the different absorption and transmission rate of X-rays in different tissues of body. It is an effective way to find the body's lesions by measuring human body with highly sensitive instruments. The detection of pulmonary nodules through observing CT images, and then the discrimination of pulmonary nodules characteristics is the main and effective method to determine benign and malignant pulmonary nodules [4]. Due to the complex structure and various shapes of the human lung tissues, it is more difficult to recognize the isolated pulmonary nodule characteristics by visual inspection. Spiculation and lobulation are important indicators for judging benign and malignant

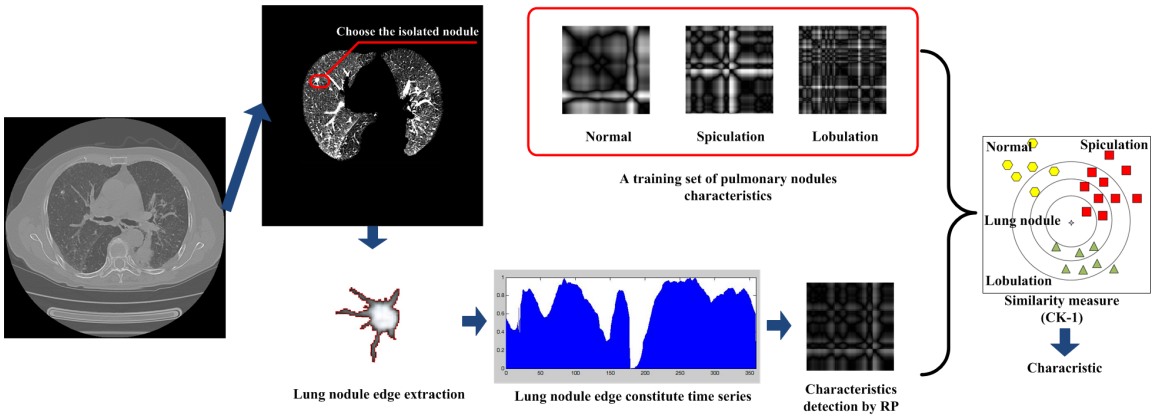


FIGURE 1. The flow chart of characteristics detection.

nodules among the numerous isolated pulmonary nodules characteristics [5], however, they have various manifestation. Therefore, the detection of isolated pulmonary nodules characteristics is the focus and difficulty of current research.

Raicu *et al.* [6] constructed a three-dimensional model and then qualitatively judged the characteristics. Horsthemke *et al.* [7] distinguished pulmonary nodules characteristics from gradient directionality. Dhara *et al.* [8] used Gaussian and mean curvature to discriminate characteristics based on differential theory. Wang *et al.* [9] enhanced the boundary of isolated pulmonary nodules image, thereby improving the physician's diagnosis correct rate. Ciompi *et al.* [10] analyzed the pulmonary nodule images from the frequency domain, and then divided the pulmonary nodules into spiculation and non-spiculation. Dhara *et al.* [11] further studied on [10] and proposed a fast discrimination algorithm. Zhang *et al.* [12] analyzed the feature of different characteristics from medical perspectives on the diagnosed pulmonary nodule characteristics data. Liu *et al.* [13] constructed classifiers based on 24 pulmonary nodule image features to distinguish pulmonary nodules characteristics. Dennie *et al.* [14] extracted boundary features by computer, and then the degree of benign and malignant of pulmonary nodules was quantitatively analyzed. Hussein *et al.* [15] constructed a CNN network to classify pulmonary nodules. Dhara *et al.* [16] established a self-learning pulmonary nodules diagnosis mechanism from the perspective of image retrieval. Hussein *et al.* [17] used multi-view neural networks to describe pulmonary nodule characteristics. Ciompi *et al.* [18] established a deep learning system from data containing pulmonary nodules to automatically extract pulmonary nodules. Xie *et al.* [19] constructed a gray level co-occurrence matrix (GLCM) based on boundary information to classify pulmonary nodules. Snoeckx *et al.* [20] researched the phenomenon of changes in nodules at different time intervals, emphasizing the importance of the pulmonary nodule edge in identifying pulmonary nodule characteristics. Li *et al.* [21] improved the Random Forest (RF) algorithm with dimensionality reduction to decrease the algorithm computation

At present, there are three key problems in distinguishing isolated pulmonary nodules. 1) Isolated pulmonary nodules are difficult to determine and extract the boundary. 2) The more image features used, the importance of the boundary information in distinguishing pulmonary nodule characteristics are weakening. 3) There are many parameters settings and the discriminant effect of characteristics cannot be displayed intuitively. All of the above are different from the physical's detection process. Accordingly, we conducted a study based on the physical's process of distinguishing pulmonary nodules characteristics, focusing on the boundary information, and using the principle of similarity to intuitively display and distinguish pulmonary nodules characteristics.

II. LUNG NODULES CHARACTERISTICS DETECTION

Through the further study of 10 professional physicals detection process of pulmonary nodules characteristic. It was found that the physical's detection process was divided into the following three steps:

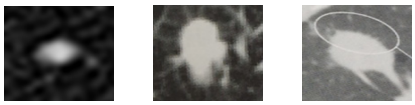
1) Focusing on the lung area and locating the isolated lung nodule area;
2) Focusing on the pulmonary nodule edge area;
3) Using the knowledge of medicine, anatomy, etc. to construct a pulmonary nodule model;
4) Making a diagnosis by using experience to distinguish the similarity of the pulmonary nodule characteristics to the known.

This paper proposes an algorithm for the computer simulation to physically distinguish the pulmonary nodule characteristics as Fig. 1. 1) Pulmonary parenchymal extraction. The physician manually determines the pulmonary nodules area. 2) Pulmonary nodules boundary information are completely extracted, and then analyzed in a time-series manner according to the strong correlation of the boundary. 3) A recurrence plot model is established for known characteristics data and discriminating characteristics data, then the

boundary information is visually displayed. 4) The similarity degree between video frames are introduced to distinguish pulmonary nodule characteristics.

A. ISOLATED PULMONARY NODULE EXTRACTION

Histopathological analysis showed that isolated pulmonary nodules were distributed in isolated areas with high local luminance. The image is shown as an area with high grayscale value and limited pixel aggregation in the lung [2]. The benign pulmonary nodules are circular with smooth and clear boundary. Malignant pulmonary nodules are quasi-circular and have vague boundary, spiculation and lobulation are the key feature to distinguish the benign and malignant pulmonary nodules. External characteristics of pulmonary nodules are shown in Fig. 2.



(a) smooth (b) lobular (c) spicule

FIGURE 2. External characteristics of pulmonary nodules.

Currently, technologies extract and detect pulmonary nodules based on computer-assisted can be divided into the completely manual method, the semi-automatic interactive method, and the automatic detection method [22]. Completely, manual method requires specialist physicians to spend lots of effort on labeling pulmonary nodules but has low efficiency. Automated detection method, without the participation of doctors, directly extracts pulmonary nodules in areas with high efficiency, but it cannot guarantee 100% accurate results. Taking into account the need for efficiency and accuracy, we use a semi-automatic interactive method for extracting pulmonary nodules.

The pixel value of the image has a high dynamic range (0-65535) according to the principle of the CT image, so it is necessary to constantly adjust the window width and window location in order to observe the entire CT image. The effect is shown in Fig. 3.

FIGURE 3. The display results of CT image in different window width and location.

For reducing the amount of data and the dynamic range of images. Moreover, according to lung nodules distributed in the lung parenchymal region and the characteristics of the lung CT data in concordance with the double-Gaussian distribution. This paper uses the double-Gaussian distribution algorithm proposed in the literature [23], which to determine the threshold and completely extract the pulmonary region. Then, the physician manually labeled the approximate location of the pulmonary nodule and the computer took a complete extraction of the pulmonary nodule in the subsequent.

The pulmonary nodule boundary information is the most important feature for detecting pulmonary nodule characteristics, hence the complete extraction of the pulmonary nodule boundary information is particularly important. Robert operator [24] fund edges by calculating local difference operators, which is accurate but sensitive to noise and not smooth enough. The Prewitt [25] and Sobel [26] operators incorporated pixel position information can detect grayscale-graded, low-noise images, but processing images with mixed, complex noises does not perform well.

In order to meet the needs of precise segmentation, scholars proposed segmentation algorithm based on level set, such as Snake [27], GAC [28], C_V [29] model and so on. They have achieved a good segmentation effect for the obvious boundary and the smoothness situation. However, when the boundary bump changes greatly, the model tends to fall into a local minimum value in the process of minimizing the energy functional, resulting in unsatisfactory segmentation effect. Accordingly, based on GAC and C_V model, this paper proposes an algorithm based on the driving force of image region information.

The GAC model is based on image gradient modulus algorithm to minimize energy functional to determine dynamic contours:

$$E(C) = \int_0^{L(C)} g(|\nabla I[C(s)]|)ds \tag{1}$$

The gradient descent function is:

$$\frac{\partial C}{\partial t} = g(|\nabla I|)k\mathbf{N} + (\nabla g \bullet \mathbf{N})\mathbf{N} \tag{2}$$

where $C(s)$ is the evolution curve, is the distance function, and k is the curvature. $\boldsymbol{N}$ is the unit normal vector of the closed curve and the direction is directed to the interior of the curve.

It can be seen that when the image is in the flat region $|\nabla \boldsymbol{I}| \approx 0$, then $g = 1$ and $\nabla g = 0$. The curve evolves by $k\boldsymbol{N}$. When $g = 0$ near the target edge, the curve evolves according to $(\nabla\, \mathrm{g} \cdot \boldsymbol{N})\boldsymbol{N}$. ∇g points in the direction of the border, that is, toward the border.

For targets with inconspicuous boundaries, the GAC model is difficult to segment effectively. Chan and Vese [30] proposed an energy functional:

$$\begin{aligned} E(c_1, c_2, C) = {} & \mu \iint_{\Omega} \delta(u)|\nabla u|\,dxdy \\ & + \lambda_1 \iint_{\Omega_1} (I - c_1)^2 H(u)\,dxdy \\ & + \lambda_2 \iint_{\Omega_2} (I - c_2)^2 (1 - H(u))\,dxdy \end{aligned} \tag{3}$$

$$c_1 = \frac{\iint_{\Omega} I \bullet H(u)\,dxdy}{\iint_{\Omega} H(u)\,dxdy}$$

$$c_2 = \frac{\iint\limits_{\Omega} I \bullet (1 - H(u))\, dxdy}{\iint\limits_{\Omega} (1 - H(u))\, dxdy} \tag{4}$$

$$H(u) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{u}{\varepsilon}\right)\right) \tag{5}$$

The image is divided into Ω_1 and Ω_2 by the curve C. When the curve reaches the vicinity of the contour line, the corresponding energy of Ω_1 and Ω_2 is 0, and $E(c_1, c_2, C)$ reaches the minimum value.

When the depression is deep, the curve expands outward in the depression of above algorithm resulting in the curve has difficult in close to the edge, which lead to inaccurate segmentation. Functionality is apt to falling into local minimums that severely affect the identification of spiculation.

It is hard to avoid this problem from the image boundary information. For this reason, we propose a driving function based on statistical information of the image region to determine the evolution of the curve.

$$D(I) = sgn(c_1 - c_2)\frac{I - \sqrt{c_1 c_2}}{\max\left(\left|I - \sqrt{c_1 c_2}\right|\right)} \tag{6}$$

The descending flow function corresponding to the GAV principle is:

$$\frac{\partial u}{\partial t} = D|\nabla u|\left(div\left(\frac{\nabla u}{|\nabla u|}\right) + \alpha\right) \tag{7}$$

where $|u| = 1$, α is the force coefficient, then

$$\frac{\partial u}{\partial t} = \alpha D|\nabla u| \tag{8}$$

As shown in Fig. 4, there are four main positions for the initial contour and the target contour. The gray is the initial contour, the white is the target contour, the blue is the area we need to discuss, and the arrows is the unfold direction of the contour. For convenience of explanation, let:

$$f(x, y) = I(x, y) - \sqrt{c_1 c_2} \tag{9}$$

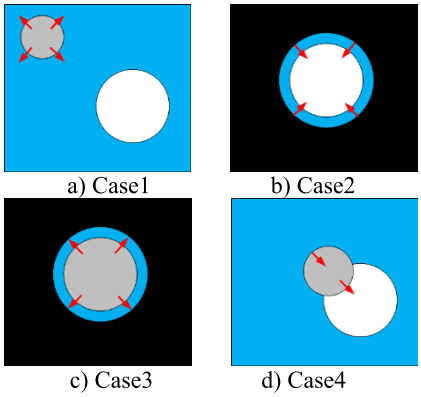


FIGURE 4. Initial contour and target contour position.

The deduction of algorithm as shown in Table 1. "+" means initial contour expansion and "-" means initial contour contraction.

TABLE 1. The deduction of algorithm.

Case	Initial state	Discussion area (blue)	D(*I*)
1	$c_1 < c_2$	$f(x, y) < 0$	+
2	$c_1 > c_2$	$f(x, y) < 0$	-
3	$c_1 > c_2$	$f(x, y) > 0$	+
4	$c_1 > c_2$	$f(x, y) < 0$	-
	$c_1 < c_2$	$f(x, y) > 0$	+

According to the above analysis, the algorithm takes the image region statistical information as the driving function, which overcomes the bad case when the boundary is vague and the depression is deep.

B. OTHER RECOMMENDATIONS BOUNDARY EXPANSION

The boundary is the most important feature for judging the signs of pulmonary nodules [20]. In order to observe the pulmonary nodule boundary, the lung nodules are unfolded at first, and then the sequence is constructed with the dependence of the boundary to prepare for the subsequent characteristics detection.

After acquiring the edge image of the pulmonary nodule, focusing on the edge area and unfolding it, the step as follows:

STEP 1: Calculate the center of the pulmonary nodule. A Cartesian coordinates is established taking the center of the pulmonary nodule as the origin. Fig. 5 is the coordinate system diagram based on the center of the pulmonary nodule, the O point represents the center point of the pulmonary nodule, and the x-axis and the y-axis are Cartesian coordinate system centered on O.

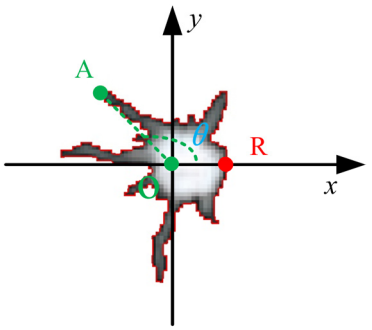


FIGURE 5. Schematic diagram of the unfolding of pulmonary nodules.

STEP 2: Unfold the pulmonary nodule. In the angle range of 0∼360°, taking the 0° position in the boundary sequence BL as the starting point (denoted as R), the angle and distance between the BL and the central point of the pulmonary nodule is calculated point by point in counterclockwise tracking. In order to ensure the distances are all positive, it is necessary to solve the four quadrants with different inverse trigonometric functions. As shown in the Fig. 5, the coordinates of point A are (x_a, y_a).

The calculation formula for θ is:

$$\theta = \begin{cases} \arccos\left(\dfrac{x_a}{\sqrt{x_a^2+y_a^2}}\right) \times \dfrac{180}{\pi} x_a \geq 0; y_a \geq 0 \\ 180 - \arcsin\left(\dfrac{y_a}{\sqrt{x_a^2+y_a^2}}\right) \times \dfrac{180}{\pi} x_a < 0; y_a \geq 0 \\ 180 + \arctan\left(\dfrac{y_a}{\sqrt{x_a^2+y_a^2}}\right) \times \dfrac{180}{\pi} x_a < 0; y_a < 0 \\ 360 - \arccos\left(\dfrac{y_a}{\sqrt{x_a^2+y_a^2}}\right) \times \dfrac{180}{\pi} x_a \geq 0; y_a < 0 \end{cases} \tag{10}$$

STEP 3: The pulmonary nodule boundary tracking sequence BL was reordered from 0° to 360° angle, and normalized to obtain a sequence CL of the pulmonary nodule edge shape.

After unfolding the pulmonary nodule, the fluctuation of the boundary can be visually reflected, but it should be measured by selecting an appropriate analytical method.

Traditional statistical analysis methods are based on the assumption that data sequences have independence. However, time series analysis [31], [32] is different from traditional statistical methods, which focuses on researching and analyzing the interdependence of data sequences. This method has been successfully applied to speech signal processing and has achieved good research results.

Similar to the speech signal, the pulmonary nodule boundary has features of continuity and strong dependence of adjacent pixels. For this reason, this paper proposes to convert the pulmonary nodule boundary sequence into a time series.

The time series is a sequence uniformly marked by time, but the value of pulmonary nodules angle regarded as ''time'' is sparse and discrete. Therefore, we need to interpolate and uniformly sample unfolding sequence CL in order to convert the pulmonary nodule edge CL into time series.

In this paper, the current mature cubic spline function is used to interpolate, and then the time series is formed by sampling. Thus, pulmonary nodules are converted to a time series DL for subsequent characteristics detection.

C. ESTABLISH RECURRENCE PLOT MODEL

The physician discrimination of pulmonary nodules is mainly through anatomical, pathological, imaging and combined with their own experience to form a reference model of different characteristics. Then, the pulmonary nodule characteristics to be identified are modeled. Finally, the similarity criteria are established to compare the discriminated model with the reference model. Computer simulation of physician detection process focuses on DL at first. Then, a reference model based on recurrence plot is constructed according to known characteristics data sets. Finally, the normalized compression distance is used to quantify the similarity degree to identify the pulmonary nodule characteristics.

The researches on time series similarity discrimination mainly include based on dynamic time warping (DTW) algorithm [33]. The DTW algorithm calculates the time axis curvature to obtain the minimum distance between two time series, and determines the best correspondence of each point in a time series, which can effectively solve other algorithms problems such as Euclidean distance is very difficult to deal with.

However, for the shape recognition of pulmonary nodules, the edge of the pulmonary nodules is unfolding from different angles, which will produce different time series, but the DTW algorithm can only search for similar areas nearby. Therefore, DTW show not excellent recognition performance when faced with this problem, and without the feature of strong similarity within the group and weak similarity between groups.

To this end, we propose to use the recurrence plot (RP) [34] method to study the boundary sequence of pulmonary nodules, so as to make full use of the chaotic, non-stationary and periodicity of the boundary, and better represent the internal structure of the sequence.

$$R_{\mathrm{i,j}} = sgn(\varepsilon - \|\vec{x}(i) - \vec{x}(j)\|), \vec{x}(\cdot) \in R^m, i, j = 1...N \tag{11}$$

where N is the number of states experienced by the time series, m is the sequence dimension, $\vec{x}(i)$ and $\vec{x}(j)$ is the value observed in the sequence at i and j, $\|\cdot\|$ is the distance between the two observation points, and ε is the threshold for measuring the difference in distance.

The recurrence plot can reflect the internal structure of the sequence, it needs to compare the similarity between sequences based on recurrence plot subsequently.

D. DISCRIMAINATION MODEL

At present, the distance algorithm for evaluating the similarity of two images is the normalized compression distance (NCD), which needs to linearize the two images, resulting in loss of spatial position information, CK-1 [36] used the inter-frame compression technique of video to extend the limitations of the NCD algorithm, which used the MPEG-1 video compression algorithm to preserve spatial information and find patterns repetitive appear in the framework to achieve compression. Compared the size of the compressed image to measure the distance between images. Accordingly, we use CK-1 to calculate the distance, the formula is as follows:

$$\mathrm{d}_{mpeg}(a, b) = \frac{C(a\,|b) + C(b\,|a)}{C(a\,|a) + C(b\,|b)} - 1 \tag{12}$$

where a is a recurrence plot that needs to determine pulmonary nodules characteristics, and b is a recurrence plot of known pulmonary nodules characteristics, $C(b|a)$ represents the video size by compressing b first and then compressing a through the MPEG-1 algorithm. The smaller $\mathrm{d}_{mpeg}(a,b)$, the higher the similarity.

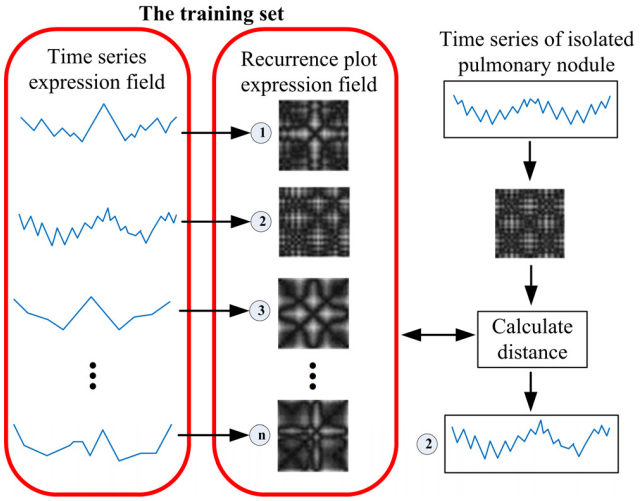


FIGURE 6. Schematic diagram of detecting lung morphological signs by PRCD algorithm.

The schematic diagram of PRCD algorithm for morphological characteristics detection of lung is shown in Fig. 6. First, different types of pulmonary nodules are expanded into time series, and then time series is transformed into Recurrence plot to form a pulmonary nodule sign database. When the solitary pulmonary nodule sign is judged, the pulmonary nodule is also transformed into a Recurrence plot. The similarity between the pulmonary nodule sign and the characteristics of the pulmonary nodule pool is judged by Eq.12, and then the pulmonary nodule sign is determined.

III. EXPERIMENT

The experiment data is provided by the Lung Image Database Consortium (LIDC) published by the National Cancer Institute (NCI), which contains the patient's complete lung CT image sequence, and the medical characteristics of the pulmonary nodules marked by four experts and the label of benign and malignant diagnosis of nodules. We selected 35 definite cases with isolated nodules. The label of pulmonary nodule characteristics by expert as a diagnostic reference, and then we establish a database for subsequent experiments.

A. SEGMENTATION PERFORMANCE

The area overlap measure (AOM) was adopted as the evaluation index of segmentation effect which is defined as:

$$\mathrm{AOM}(\vartheta, \xi) = \frac{\mathrm{S}(\vartheta \cap \xi)}{\mathrm{S}(\vartheta \cup \xi)} \times 100\% \tag{13}$$

where AOM is the area overlap measure, ϑ is the image marked by the physical, ξ is the segmentation result graph, $S(\cdot)$ indicates the number of pixels in the corresponding area, and the larger the AOM value, the better the segmentation effect.

Compared the typical algorithm and the mainstream algorithm with our proposed algorithm, the results of the AOM average are shown in Table 2.

TABLE 2. Comparison of segmentation results.

Algorithm	AOM
Robert[24]	65%
Prewitt[25]	71%
Snake[27]	83%
GAC[28]	85%
C_V[29]	87%
Ours	95%

The Robert [24] and Prewitt [25] algorithms have good segmentation results for benign pulmonary nodules, but the malignant nodules are poorly segmented due to vague boundaries. The algorithm [27], [28], [29] based on level set is better than the classical algorithm. The Snake [27] algorithm uses gradient and curvature constraints has good segmentation in the better rule boundaries. The GAC [28] algorithm constructs an internal and external force model to segment the image. C_V [29] algorithm has improved the segmentation effect of pulmonary nodules with vague boundaries due to the constraints of the model. The level set algorithm proposed above is easy to fall into local minimum value when the boundary is vague and the shape is concave and convex, resulting in the segmentation effect is not good. The algorithm proposed in this paper can use the local image region statistical information to completely segment the lung nodules.

B. COMPARISION OF SIGNS DISCRIMINANT ALGORITHMS

The discriminant results were measured by four indicators: accuracy, sensitivity, specificity and ROC (receiver operating characteristic curve) curve. We introduce the following concepts at first:

True positive (*TP*): Positive samples are predicted to be positive

False positive (*FP*): negative samples are predicted to be positive

True negative (*TN*): negative samples are predicted to be negative

False negative (*FN*): positive samples are predicted to be negative

Accuracy (*ACC*) is used to reflect the correct diagnostic example for all instance ratios. Sensitivity (*SEN*) reflects the diagnostic performance of positive examples. Specificity (*SPE*) response to the diagnostic performance of negative examples.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

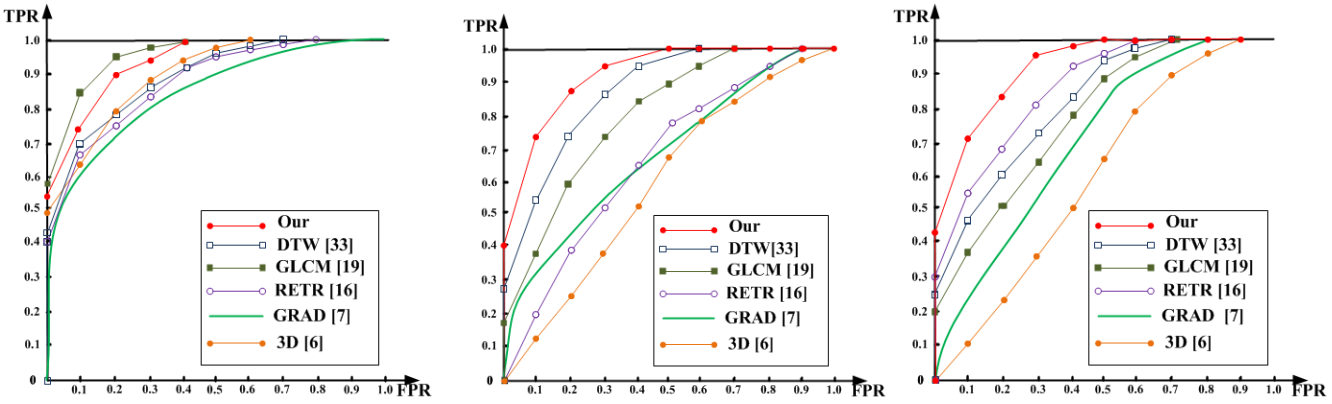
$$SEN = \frac{TP}{TP + FN} \tag{15}$$

$$SPE = \frac{TN}{TN + FP} \tag{16}$$

The ROC curve can be used to evaluate the performance of two or more image classification algorithms, and

TABLE 3. Pulmonary nodule characteristics discriminant results.

Algorithm	Normal			Spiculation			Lobulation			Average processing time (s/nodule)
	ACC%	*SEN*%	*SPE*%	*ACC*%	*SEN*%	*SPE*%	*ACC*%	*SEN*%	*SPE*%	
3D [6]	75.6	73.5	82.1	73.6	71.6	77.6	73.5	76.5	74.3	1.2
GRAD [7]	74.1	70.3	81.9	74.3	72.7	83.3	77.1	77.8	76.5	0.5
RETR [16]	76.3	71.5	79.6	76.5	77.5	85.6	83.6	90.1	87.4	0.8
GLCM [19]	**80.6**	**81.4**	**83.2**	75.4	79.5	86.7	82.4	87.6	74.5	0.63
DTW [33]	77.4	76.2	74.3	85.7	83.8	88.3	82.9	88.9	76.5	**0.55**
Ours	78.9	79.4	82.4	**88.6**	**85.9**	**89.6**	**89.4**	**91.4**	**88.2**	0.58



a) Normal pulmonary nodule b) Spiculation pulmonary nodule c) Lobulation pulmonary nodule

FIGURE 7. ROC curve.

it has extremely wide applications in clinical and scientific research.

From the analysis of Table 3, it can be seen that the boundary of the normal pulmonary nodule is smooth, and the detection results of all algorithms are not much different. In the face of spiculation and lobulation, the results are quite different because of the diversity of their features. 3D algorithm [6] analyzed pulmonary nodule characteristics in three-dimensional angle, the algorithm has high algorithm complexity and slow processing time, and it did not establish a unified three-dimensional detection model. GRAD (gradient) algorithm [7] used gradient directionality to discriminate pulmonary nodule characteristics, which has faster speed. RETR (retrieval) algorithm [16] established a retrieval model from the aspect of image information to identify pulmonary nodule characteristics, which has certain discriminating effects. And GLCM (gray level co-occurrence matrix) algorithm [19] fused multiple features to distinguish pulmonary nodule characteristics.

The DTW algorithm [33] focuses on the boundary information with the fast processing speed. The difference of the boundary sequences formed from different positions is not considered resulting in insufficient detection effect. It can be seen from the ROC curve of Fig. 7, although the algorithm proposed in this paper has lower effect on the normal nodule than the GLCM algorithm, it has better results in the face of the lobulation and speculation, and the time meets the clinical application criteria.

In this paper, we can conclude from the algorithm that the accurate segmentation of pulmonary nodules is the premise of the realization of pulmonary nodule signs, so it takes more time to establish the pulmonary nodule extraction model. Generally, although 0.58s/nodule is slightly higher than 0.55s/nodule of DTW, the detection accuracy is greatly improved, which can meet the actual needs.

C. ALGORITHM PROCESS

The proposed algorithm discriminates different types of pulmonary nodules, the effects of each step are shown in Fig. 8. Red is the boundary extracted by the algorithm. Different pulmonary nodules are unfolded into time series, so the difference between normal and lobulation with speculation could be easily distinguished by comparing the smoothness of time series. However, the time series of the normal and lobulation are smooth cannot be distinguished effectively. The recurrence plot algorithm comprehensively considers the chaotic, non-stationary and periodicity of the pulmonary nodule boundary, therefore the difference between normal and lobulation is obvious.

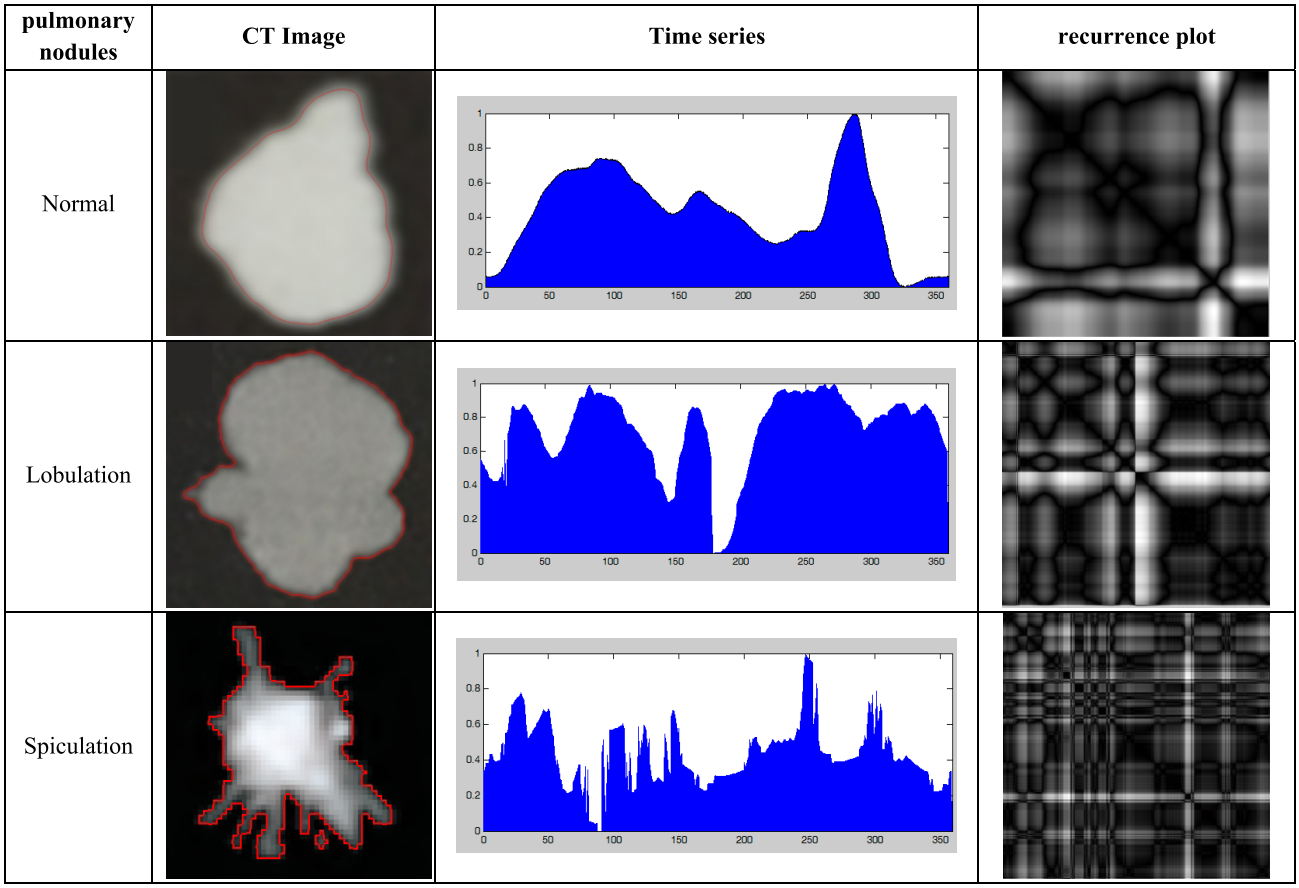


FIGURE 8. The characteristics detection result image.

IV. CONCLUSION

In order to reduce the intensity of doctors' work and improve the efficiency of doctors' sign discrimination, a complete process of pulmonary nodule extraction and sign recognition is proposed according to the problems of difficulties in segmentation and recognition of pulmonary nodule by computer. The computer is used to simulate the detection process of doctors, and the accurate segmentation of pulmonary nodules is proposed based on the statistical information of image regions. Then we focus on the pulmonary nodule boundary information to expand the pulmonary nodule into a time series. Finally, its boundary information is intuitively displaced through the establishment of recursive graph and its system is discriminated through the distance of video similarity. The recognition of pulmonary nodule signs is realized, which shows a good effect on the existing database. In the future, we will study the benign and malignant pulmonary nodules on this basis.

REFERENCES

[1] H. B. Heidinger, K. R. Anderson, E. M. Moriarty, D. B. Costa, S. P. Gangadharan, P. A. VanderLaan, and A. A. Bankier, "Size measurement and T-staging of lung adenocarcinomas manifesting as solid nodules ≤30 mm on CT: Radiology-pathology correlation," *Academic Radiol.*, vol. 24, no. 7, pp. 851–859, 2017.

[2] G. Pérez and P. Arbeláez, "Automated detection of lung nodules with three-dimensional convolutional neural networks," *Proc. SPIE*, vol. 10572, Nov. 2017, Art. no. 1057218.

[3] G. A. Silvestri, N. T. Tanner, P. Kearney, A. Vachani, P. P. Massion, A. Porter, S. C. Springmeyer, K. C. Fang, D. Midthun, and P. J. Mazzone, "Assessment of plasma proteomics Biomarker's ability to distinguish benign from malignant lung nodules: Results of the PANOPTIC (pulmonary nodule plasma proteomic classifier) trial," *Chest*, vol. 154, no. 3, pp. 491–500, 2018.

[4] E. R. Miller, R. K. Putman, M. Vivero, Y. Hung, T. Araki, M. Nishino, H. Hatabu, L. Sholl, and G. M. Hunninghake, "Interstitial lung abnormalities and histopathologic correlates in patients undergoing lung nodule resection," in *Proc. IPF, Clin. Stud., Therapeutics*, 2017, p. A1120.

[5] E. Paci, D. Puliti, A. L. Pegna, L. Carrozzi, G. Picozzi, F. Falaschi, F. Pistelli, F. Aquilini, M. Zappa, F. Carozzi, and M. Mascalchi, "MA01. 09 mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial (ITALY)," *J. Thoracic Oncol.*, vol. 12, no. 1, pp. S346–S347, 2017.

[6] D. S. Raicu, E. Varutbangkul, J. G. Cisneros, J. D. Furst, D. S. Channin, and S. G. Armato, "Semantics and image content integration for pulmonary nodule interpretation in thoracic computed tomography," *Proc. SPIE*, vol. 6512, Mar. 2007, Art. no. 65120S.

[7] W. H. Horsthemke, D. S. Raicu, and J. D. Furst, "Characterizing pulmonary nodule shape using a boundary-region approach," *Proc. SPIE*, vol. 7260, Feb. 2009, Art. no. 72602Y.

[8] A. K. Dhara, S. Mukhopadhyay, N. Alam, and N. Khandelwal, "Measurement of spiculation index in 3D for solitary pulmonary nodules in volumetric lung CT images," *Proc. SPIE*, vol. 8670, Feb. 2013, Art. no. 86700K.

[9] Y.-X. J. Wang, J. S. Gong, K. Suzuki, and S. K. Morcos, "Evidence based imaging strategies for solitary pulmonary nodule," *J. Thoracic Disease*, vol. 6, no. 7, pp. 872–887, 2014.

[10] F. Ciompi, C. Jacobs, E. T. Scholten, S. J. van Riel, M. M. W. Wille, M. Prokop, and B. van Ginneken, "Automatic detection of spiculation of pulmonary nodules in computed tomography images," *Proc. SPIE*, vol. 9414, Mar. 2015, Art. no. 941409.

[11] A. K. Dhara, S. Mukhopadhyay, P. Saha, M. Garg, and N. Khandelwal, "Differential geometry-based techniques for characterization of boundary roughness of pulmonary nodules in CT images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 3, pp. 337–349, 2016.

[12] Y. Zhang, Y. Shen, J. W. Qiang, J. D. Ye, J. Zhang, and R. Y. Zhao, "HRCT features distinguishing pre-invasive from invasive pulmonary adenocarcinomas appearing as ground-glass nodules," *Eur. Radiol.*, vol. 26, no. 9, pp. 2921–2928, 2016.

[13] Y. Liu, Y. Balagurunathan, T. Atwater, S. Antic, Q. Li, R. C. Walker, G. T. Smith, P. P. Massion, M. B. Schabath, and R. J. Gillies, "Radiological image traits predictive of cancer status in pulmonary nodules," *Clin. Cancer Res.*, vol. 23, no. 6, pp. 1442–1449, 2016.

[14] C. Dennie, R. Thornhill, V. Sethi-Virmani, C. A. Souza, H. Bayanati, A. Gupta, and D. Maziak, "Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules," *Quant. Imag. Med. Surg.*, vol. 6, no. 1, pp. 6–15, 2016.

[15] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using $_3$D CNN-based multi-task learning," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 249–260.

[16] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "Content-based image retrieval system for pulmonary nodules: Assisting radiologists in self-learning and diagnosis of lung cancer," *J. Digit. Imag.*, vol. 30, no. 1, pp. 63–77, 2017.

[17] S. Hussein, R. Gillies, K. Cao, Q. Song, and U. Bagci, "TumorNet: Lung nodule characterization using multi-view convolutional neural network with Gaussian process," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 1007–1010.

[18] F. Ciompi, K. Chung, S. J. Van Riel, A. Arindra A. Setio, P. K. Gerke, C. Jacobs, E. Th. Scholten, C. Schaefer-Prokop, M. M. W. Wille, A. Marchianò, U. Pastorino, M. Prokop, and B. van Ginneken, "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," *Sci. Rep.*, vol. 7, Apr. 2017, Art. no. 046479.

[19] Y. Xie, Y. Xia, J. Zhang, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Inf. Fusion*, vol. 42, pp. 102–110, Jul. 2018.

[20] A. Snoeckx, P. Reyntiens, D. Desbuquoit, M. J. Spinhoven, P. E. Van Schil, J. P. van Meerbeeck, and P. M. Parizel, "Evaluation of the solitary pulmonary nodule: Size matters, but do not ignore the power of morphology," *Insights Imag.*, vol. 9, no. 1, pp. 73–86, 2018.

[21] X.-X. Li, B. Li, L.-F. Tian, and L. Zhang, "Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm," *IET Image Process.*, vol. 12, no. 7, pp. 1253–1264, Jul. 2018.

[22] P. Delogu, M. E. Fantacci, and I. Gori, "Computer-aided detection of pulmonary nodules in low-dose CT," in *Proc. Int. Symp. Comput. Modelling Objects Represented Images. Fundam., Methods Appl. (CompIMAGE)*, Coimbra, Portugal, Oct. 2006, p. 165.

[23] S. Qiu, D. Wen, Y. Cui, and J. Feng, "Lung nodules detection in CT images using Gestalt-based algorithm," *Chin. J. Electron.*, vol. 25, no. 4, pp. 711–718, 2016.

[24] S. K. Mishra and K. Jena, "Edge detection of images: An application of CR methodology," *Int. J. Image Process. Pattern Recognit.*, vol. 1, no. 1, 2015.

[25] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE J. Solid-State Circuits*, vol. 23, no. 2, pp. 358–367, Apr. 1988.

[26] X. Hua, "Non-local means image denoising using improved LOG operator," *J. Tongling Univ.*, vol. 4, p. 27, Apr. 2012.

[27] V. Rajinikanth, N. Madhavaraja, S. C. Satapathym, S. Chandra, and S. L. Fernandes, "Otsu's multi-thresholding and active contour snake model to segment dermoscopy images," *J. Med. Imag. Health Inform.*, vol. 7, no. 8, pp. 1837–1840, 2017.

[28] K. Zhang, L. Zhang, H. Song, and W. Zhou, "Active contours with selective local or global segmentation: A new formulation and level set method," *Image Vis. Comput.*, vol. 28, no. 4, pp. 668–676, 2010.

[29] X. Yang, W. Sun, M. Li, M. Chen, N. Ming, J. Han, W. Li, and M. Chen, "Water droplets fluorescence image segmentation of cucumber leaves based on K-means clustering with opening and closing alternately filtering," *Trans. Chin. Soc. Agricult. Eng.*, vol. 32, no. 17, pp. 136–143, 2016.

[30] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[31] M. Schultzberg, B. Muthén, T. Asparouhov, and E. Hamaker, "Dynamic structural equation modeling of intensive longitudinal data using multilevel time series analysis in Mplus version 8 part 7," in *Proc. Workshop Johns Hopkins Univ.*, 2017, pp. 1–48.

[32] V. Akbari, A. Doulgeris, and T. Eltoft, "Time series analysis of multi-polarisation synthetic aperture radar images with a textural-contextual model," in *Proc. 9th Eur. Conf. Synth. Aperture Radar*, Apr. 2012, pp. 726–729.

[33] S. Soheily-Khah and P.-F. Marteau, "Sparsification of the alignment path search space in dynamic time warping," 2017, *arXiv:1711.04453*. [Online]. Available: https://arxiv.org/abs/1711.04453

[34] M. Riedl, N. Marwan, and J. Kurths, "Extended generalized recurrence plot quantification of complex circular patterns," *Eur. Phys. J. B*, vol. 90, no. 3, p. 58, 2017.

[35] P. P. Vázquez and J. Marco, "Using normalized compression distance for image similarity measurement: An experimental study," *Vis. Comput.*, vol. 28, no. 11, pp. 1063–1084, 2012.

[36] B. Campana and E. J. Keogh, "A compression-based distance measure for texture," *Stat. Anal. Data Mining*, vol. 3, no. 6, pp. 381–398, 2010.

[37] G. Litjens *et al.*, "Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge," *Med. Image Anal.*, vol. 18, no. 2, pp. 359–373, 2014.

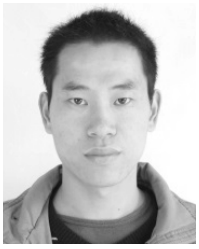
SHI QIU received the M.S. degree from Xidian University, Xi'an, China, in 2012, and the Ph.D. degree from the University of Chinese Academy of Sciences, Xi'an, China, in 2016. He is currently an Associate Research Fellow with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. His research interests include image restoration, sparse representation, and computer aided diagnosis and object detection.

QIANG GUO received the B.S. degree from the Shandong University of Technology, Zibo, China, in 2002, and the M.S. and Ph.D. degrees from Shanghai University, Shanghai, China, in 2005 and 2010, respectively. From 2012 to 2015, he was a Postdoctoral Fellow with Shandong University, Jinan, China. He is currently an Associate Professor with the School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan. His research interests include image restoration, sparse representation, and object detection.

DONGMEI ZHOU received the B.S. degree in communication engineering from Sichuan Union University (renamed as Sichuan University), China, in 1995, and the M.S. degree from the Department of Electronic Engineering, Chengdu University of Technology, China, in June 2002. She is currently an Associate Professor with the College of Information Science and Technology, Chengdu University of Technology. Her research interests include mage communication, radio frequency communication, signal processing, and traffic safety detection.

YI JIN received the Ph.D. degree in signal and information processing from the Institute of Information Science, Beijing Jiaotong University, Beijing, China, in 2010. She was a Visiting Scholar with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2013 to 2014. She is currently an Associate Professor with Beijing Jiaotong University, Beijing. Her research interests include computer vision, pattern recognition, image processing, and machine learning.

TAO ZHOU received the bachelor's degree in computer science and technology from Northwest Normal University, in 1995, the M.Eng. degree from Qiqihar University, in 2005, and the Ph.D. degree in computer science and technology from Northwest Polytechnic University, in 2010. He is currently a Professor with North Minzu University, and the Director of the Chinese Society of Image Graphics, Chinese Society of Stereology. His current research interests include medical image processing, pattern recognition, and big data analysis.



ZHEN'AN HE received the bachelor's degree in mechanical automation from the Rocket Force University of Engineering, in 2008, the M.E. degree from Xi'an Technological University, in 2012, and the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, in 2018. He is currently with the Shaanxi Institute of Medical Device Quality Supervision and Inspection. His current research interests include quality research and evaluation of medical device.